Topic Two
Regression with Time-Series Data: Nonstationary Variables

Outline

- Stationary and Nonstationary Variables
- Spurious Regressions
- Unit Root Tests for Nonstationarity
- Cointegration
- Regression When There Is No Cointegration

Stationary and Nonstationary Variables

- The time series y_t is stationary if for all values, and every time period, it is true that:

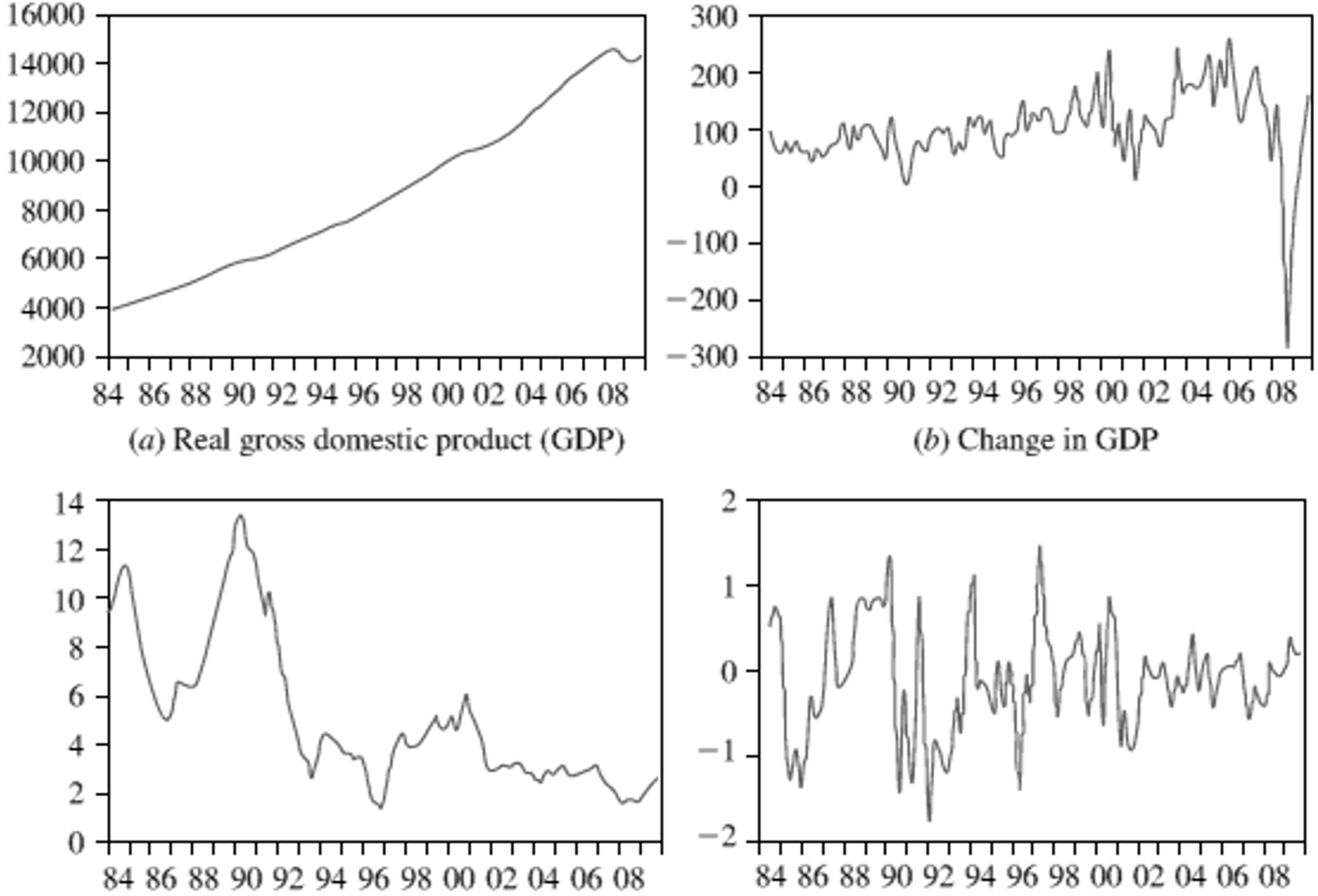
$$E(y_t) = \mu \quad \text{(constant mean)}$$

$$\text{var}(y_t) = \sigma^2 \quad \text{(constant variance)}$$

$$\text{cov}(y_t, y_{t+s}) = \text{cov}(y_t, y_{t-s}) = \gamma_s \quad \text{(covariance depends on } s\text{, not } t\text{)}$$

- Nonstationary series with nonconstant means are often described as not having the property of **mean reversion**
- Stationary series have the property of mean reversion

Stationary and Nonstationary Variables



(*a*) Real gross domestic product (GDP)

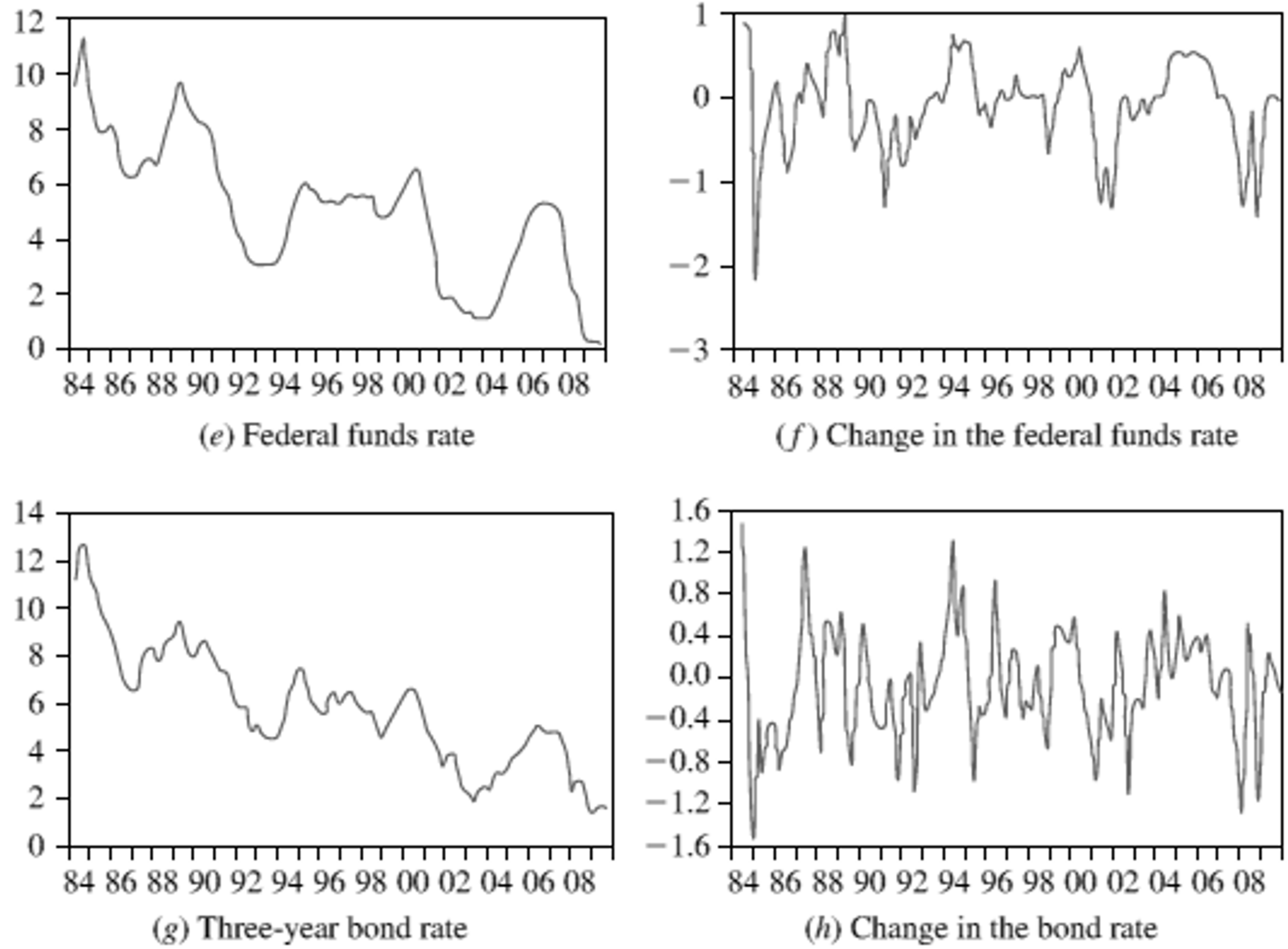
(*b*) Change in GDP

(*c*) Inflation rate

(*d*) Change in the inflation rate

SSRMC

Stationary and Nonstationary Variables



(*e*) Federal funds rate

(*f*) Change in the federal funds rate

(*g*) Three-year bond rate

(*h*) Change in the bond rate

Stationary and Nonstationary Variables

Variable	Sample periods 1984:2 to 1996:4	Sample periods 1997:1 to 2009:4
Real GDP (a)	5813.0	11458.2
Inflation rate (c)	6.9	3.2
Federal funds rate (e)	6.4	3.5
Bond rate (g)	7.3	4.0
Change in GDP (b)	82.7	120.3
Change in the inflation rate (d)	−0.16	0.02
Change in the federal funds rate (f)	−0.09	−0.10
Change in the bond rate (h)	−0.10	−0.09

Sample Means of Time Series Shown on previous slides

SSRMC

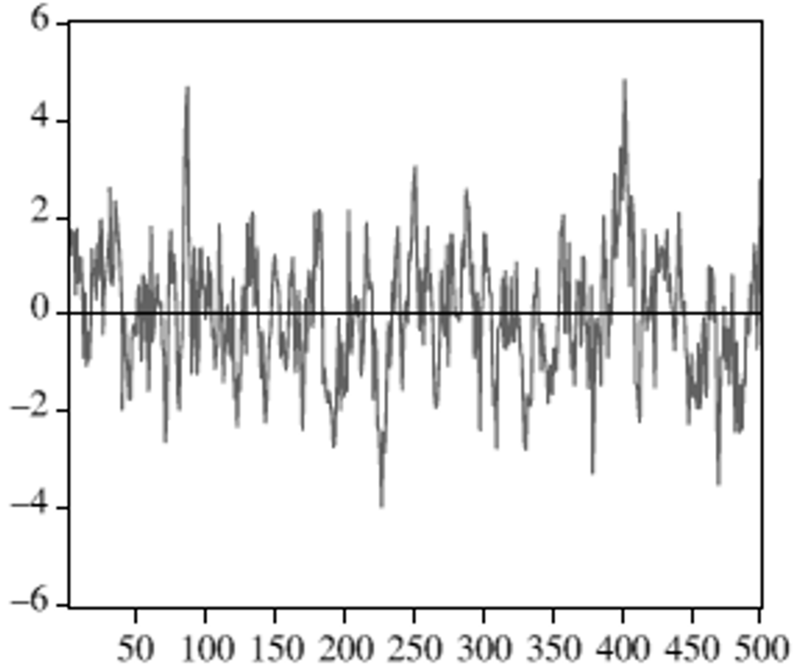
Stationary and Nonstationary Variables

- The econometric model generating a time-series variable y_t is called a **stochastic** or **random process**
- A sample of observed y_t values is called a particular **realization** of the stochastic process
- An AR(1) model

$$y_t = \rho y_{t-1} + v_t$$

  - The errors v_t are independent, with zero mean and constant variance σ_v^2, and may be normally distributed
  - The errors are sometimes known as "**shocks**" or "**innovations**"
  - It is stationary if $|\rho| < 1$

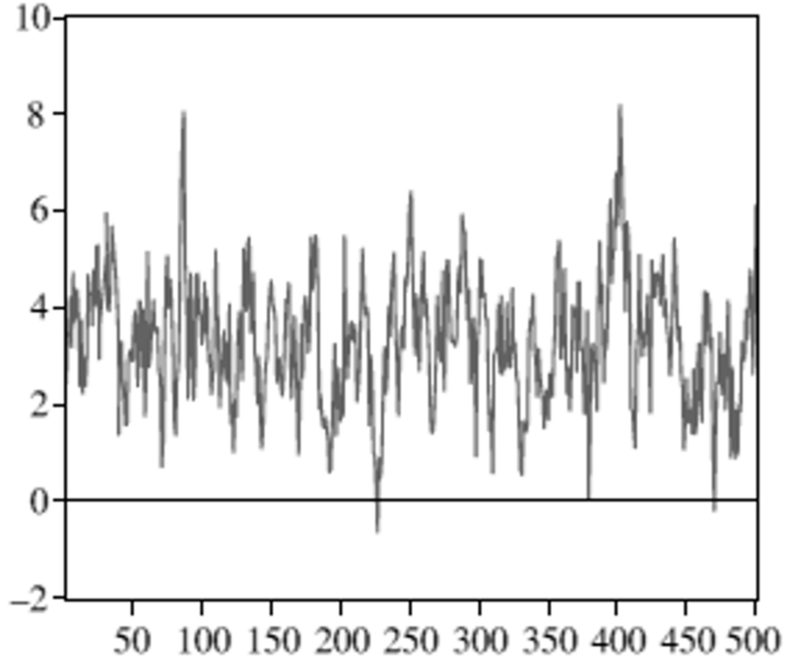
Stationary and Nonstationary Variables



$(a)\ y_t = 0.7y_{t-1} + v_t$

Stationary with a zero mean

$$E(y_t) = 0$$

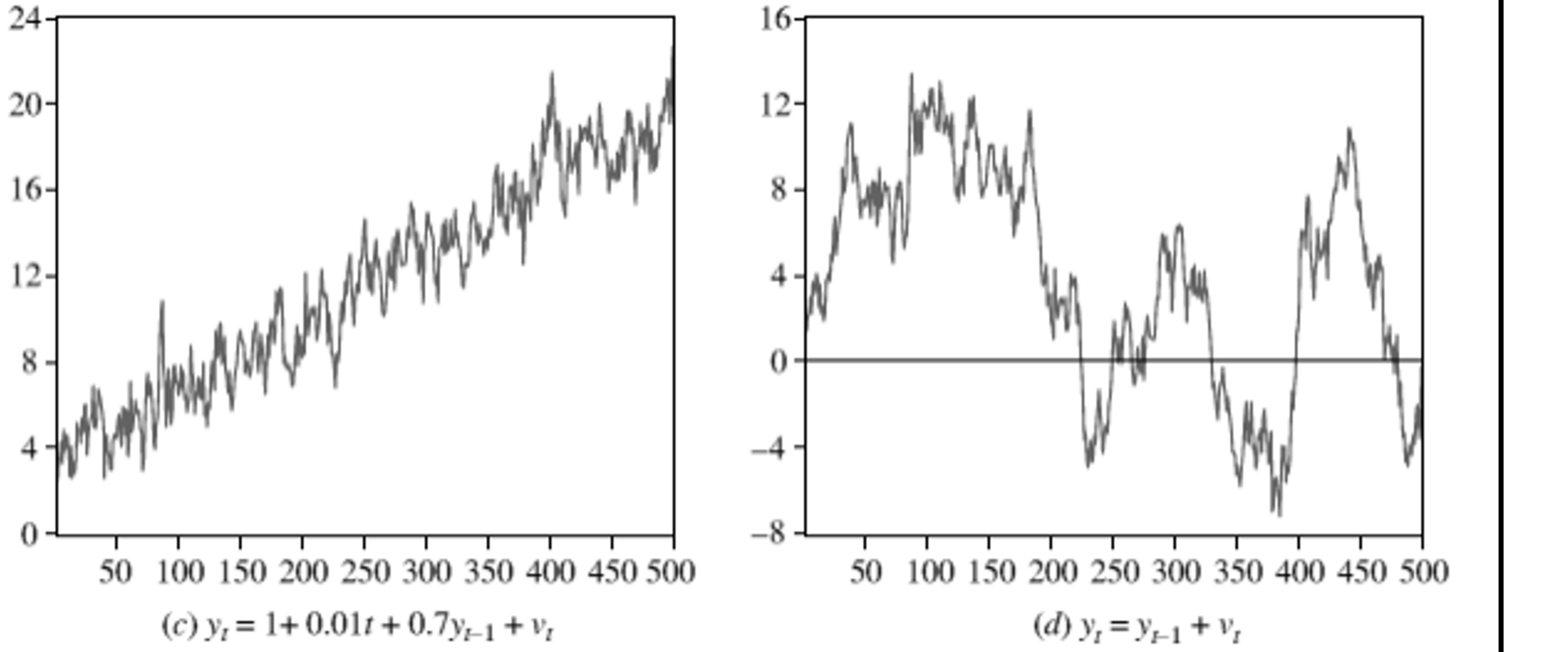


$(b)\ y_t = 1 + 0.7y_{t-1} + v_t$

Stationary with a non- zero mean

$$E(y_t) = \alpha / (1 - \rho) = 1/(1 - 0.7) = 3.33$$

Stationary and Nonstationary Variables



Stationary around
a linear trend: ($\mu + \delta t$)

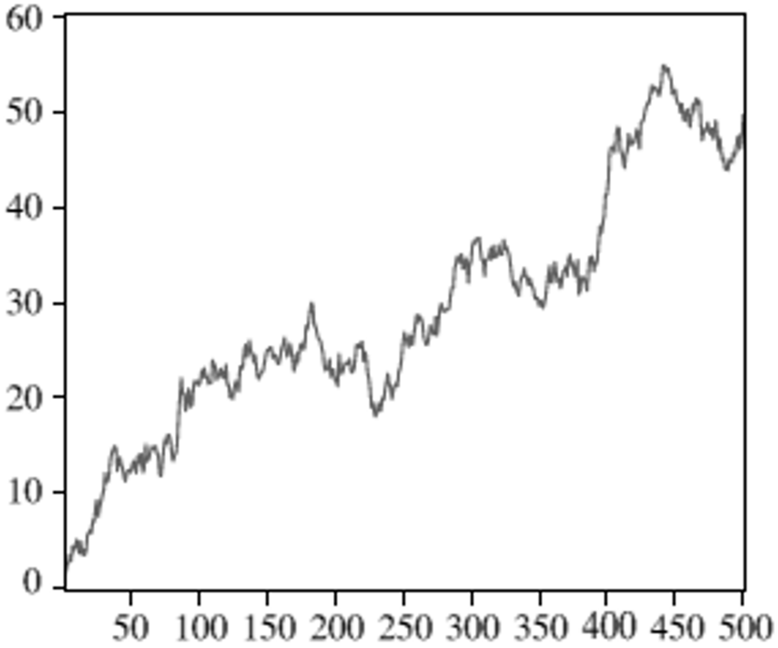
Non –stationary: Random Walk

$$y_t = y_{t-1} + v_t = y_0 + \sum_{s=1}^{t} v_s$$

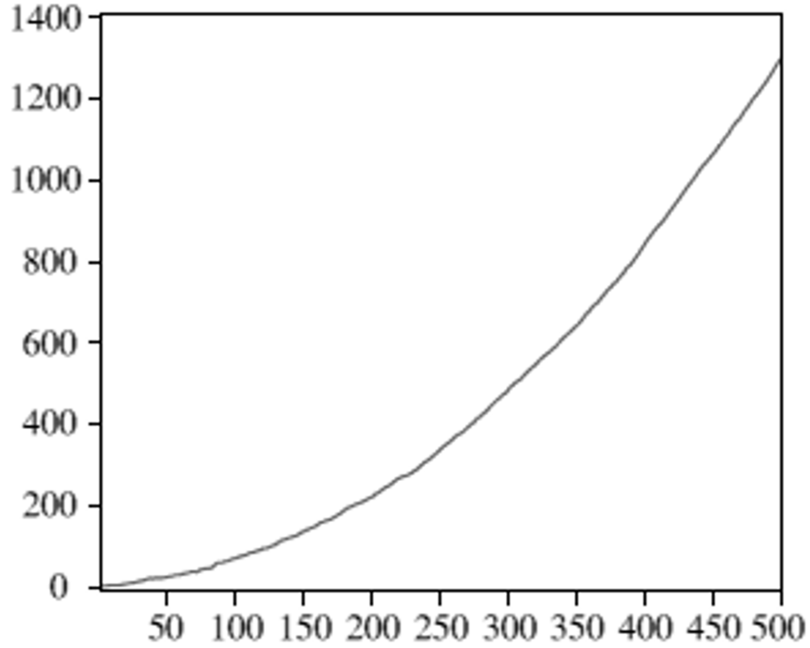
$\sum_{s=1}^{t} v_s$ is called **stochastic trend**

$$E(y_t) = y_0, \quad \text{var}(y_t) = t\sigma_v^2$$

Stationary and Nonstationary Variables



(e) $y_t = 0.1 + y_{t-1} + v_t$



(f) $y_t = 0.1 + 0.01t + y_{t-1} + v_t$

Non –stationary:
Random Walk with drift

$$y_t = \alpha + y_{t-1} + v_t = t\alpha + y_0 + \sum_{s=1}^{t} v_s$$

$t\alpha$ is a **deterministic trend**

$$E(y_t) = t\alpha + y_0, \quad \text{var}(y_t) = t\sigma_v^2$$

Non –stationary:
Random Walk with drift and trend

$$y_t = \alpha + \delta t + y_{t-1} + v_t = t\alpha + \left(\frac{t(t+1)}{2}\right)\delta + y_0 + \sum_{s=1}^{t} v_s$$

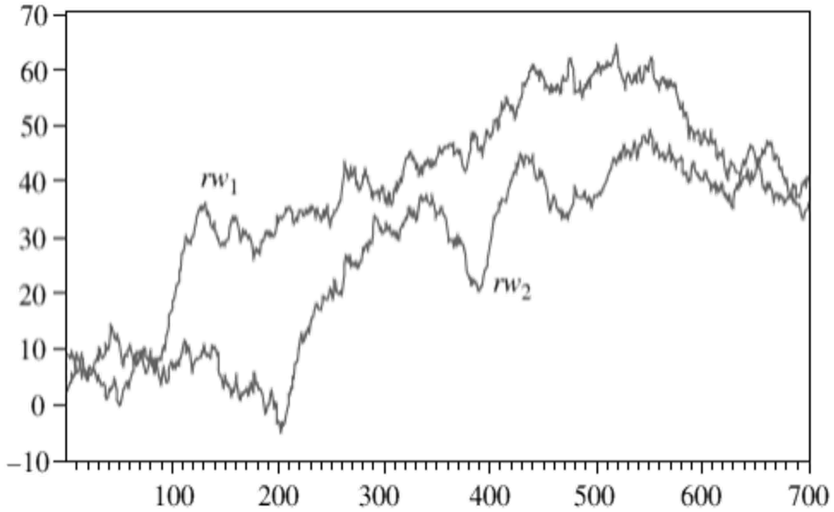
$t\alpha$ is a **deterministic trend**

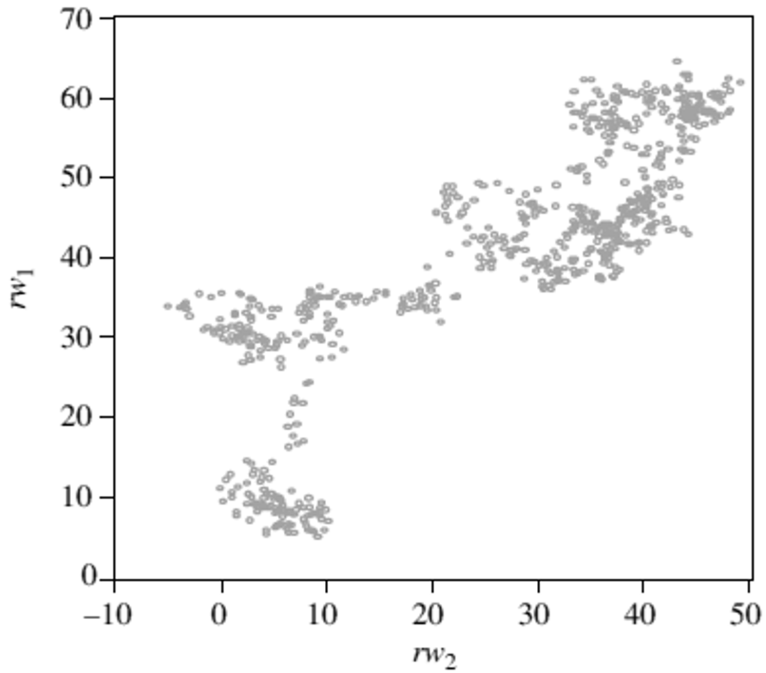
$$E(y_t) = t\alpha + \left(\frac{t(t+1)}{2}\right)\delta + y_0, \quad \text{var}(y_t) = t\sigma_v^2$$

Spurious Regressions

- Apparently significant regression results from unrelated data when nonstationary series are used in regression analysis
- Consider two independent random walks:

$$rw_1: \quad y_t = y_{t-1} + v_{1t}$$
$$rw_2: \quad x_t = x_{t-1} + v_{2t}$$





Spurious Regressions

- A simple regression of series one (rw_1) on series two (rw_2) yields:

$$\widehat{rw_{1t}} = 17.818 + 0.842\, rw_{2t}, \quad R^2 = 0.70$$
$$(t) \qquad\qquad (40.837)$$

- These results are completely meaningless, or spurious. The apparent significance of the relationship is false
- The least squares estimator and least squares predictor do not have their usual properties, and *t*-statistics are not reliable

SSRMC

Unit Root Tests for Stationarity

- The AR(1) process $y_t = \rho y_{t-1} + v_t$ is stationary when $|\rho| < 1$
- We want to test whether ρ is equal to one (nonstationary) or significantly less than one (stationary)
- Tests for this purpose are known as **unit root tests for stationarity**
- Dickey–Fuller test is the most commonly used one

SSRMC

Unit Root Tests for Stationarity

- Dickey–Fuller test
- Test equation:

$$y_t - y_{t-1} = \rho y_{t-1} - y_{t-1} + v_t$$

$$\Delta y_t = (\rho - 1) y_{t-1} + v_t = \gamma y_{t-1} + v_t$$

- Hypotheses:

$$H_0 : \rho = 1 \quad \Leftrightarrow \quad H_0 : \gamma = 0 \quad \text{vs} \quad H_1 : \rho < 1 \quad \Leftrightarrow \quad H_1 : \gamma < 0$$

- Some variations:

$$\Delta y_t = \alpha + \gamma y_{t-1} + v_t$$

$$\Delta y_t = \alpha + \gamma y_{t-1} + \lambda t + v_t$$

$$\Delta y_t = \alpha + \gamma y_{t-1} + \lambda t + \sum_{s=1}^{m} a_s \Delta y_{t-s} + v_t$$ **Augmented DF test**

SSRMC

Unit Root Tests for Stationarity

- Dickey–Fuller test
- Critical values:

Model	1%	5%	10%
$\Delta y_t = \gamma y_{t-1} + v_t$	−2.56	−1.94	−1.62
$\Delta y_t = \alpha + \gamma y_{t-1} + v_t$	−3.43	−2.86	−2.57
$\Delta y_t = \alpha + \lambda t + \gamma y_{t-1} + v_t$	−3.96	−3.41	−3.13
Standard critical values	−2.33	−1.65	−1.28

Note: These critical values are taken from R. Davidson and J. G. MacKinnon (1993), *Estimation and Inference in Econometrics*, New York: Oxford University Press, p. 708.

SSRMC

Unit Root Tests for Stationarity

- The Dickey-Fuller testing procedure:
  - First plot the time series of the variable and select a suitable Dickey-Fuller test based on a visual inspection of the plot
    - If the series appears to be wandering or fluctuating around a sample average of zero, use $\Delta y_t = \gamma y_{t-1} + v_t$
    - If the series appears to be wandering or fluctuating around a sample average which is nonzero, use $\Delta y_t = \alpha + \gamma y_{t-1} + v_t$
    - If the series appears to be wandering or fluctuating around a linear trend, use $\Delta y_t = \alpha + \gamma y_{t-1} + \lambda t + v_t$
  - Second, proceed with one of the unit root tests

Unit Root Tests for Stationarity

- The Dickey-Fuller test:
- Example:
  - The federal funds rate (F_t)
  - The three-year bond rate (B_t)
- Test equations:

$$\widehat{\Delta F_t} = 0.173 - 0.045 F_{t-1} + 0.561 \Delta F_{t-1}$$
$$(tau) \qquad (-2.505)$$
$$\widehat{\Delta B_t} = 0.237 - 0.056 B_{t-1} + 0.237 \Delta B_{t-1}$$
$$(tau) \qquad (-2.703)$$

- The 5% critical value for *tau* (τ_c) is -2.86
- Do not reject H_0. Series are not stationary.

SSRMC

Unit Root Tests for Stationarity

- The Dickey-Fuller test example (Cont'd):
- DF test on the first difference of both series

$$\widehat{\Delta(\Delta F)_t} = -0.447(\Delta F)_{t-1} \qquad \widehat{\Delta(\Delta B)_t} = -0.701(\Delta B)_{t-1}$$

$$(tau) \quad (-5.487) \qquad\qquad (tau) \quad (-7.662)$$

- The 5% critical value for *tau* (τ_c) is -1.94
- Reject H_0. The first difference series are stationary.
- Series that can be made stationary by taking the first difference, are said to be **integrated of order one**, and denoted as **I(1)**
- Stationary series are said to be integrated of order zero, **I(0)**

Cointegration

- Definition: If $e_t = y_t - \beta_1 - \beta_2 x_t$ is a stationary I(0) process, y_t and x_t are said to be **cointegrated**
- Cointegration implies that y_t and x_t share similar stochastic trends, and, since the difference e_t is stationary, they never diverge too far from each other
- In this case regression analysis is valid. No concerns about spurious regression.
- The test for stationarity of the residuals is based on the test equation: $\Delta\hat{e}_t = \gamma\hat{e}_{t-1} + v_t$
- Critical values vary according to regression model

Cointegration

- Cointegration test:
- There are three sets of critical values. Which set we use depends on whether the residuals are derived from:

$$Equation\,1: \quad \hat{e}_t = y_t - bx_t$$

$$Equation\,2: \quad \hat{e}_t = y_t - b_2 x_t - b_1$$

$$Equation\,3: \quad \hat{e}_t = y_t - b_2 x_t - b_1 - \hat{\delta} t$$

Regression model	1%	5%	10%
(1) $y_t = \beta x_t + e_t$	−3.39	−2.76	−2.45
(2) $y_t = \beta_1 + \beta_2 x_t + e_t$	−3.96	−3.37	−3.07
(3) $y_t = \beta_1 + \delta t + \beta_2 x_t + e_t$	−3.98	−3.42	−3.13

Note: These critical values are taken from J. Hamilton (1994), *Time Series Analysis*, Princeton University Press, p. 766.

Cointegration

- Cointegration test example:
- The relation between federal funds rate (F_t) and three-year bond rate (B_t)

$$\hat{B}_t = 1.140 + 0.914 F_t, \quad R^2 = 0.881$$

$$(t) \quad (6.548) \quad (29.421)$$

- The unit root test for stationarity in the estimated residuals is:

$$\Delta \hat{e}_t = -0.225 \hat{e}_{t-1} + 0.254 \Delta \hat{e}_{t-1}$$

$$(tau) \quad (-4.196)$$

- Critical value is -3.37. The two series are cointegrated. The regression results are valid.

SSRMC

Cointegration

- Error Correction Model
- Consider an autoregressive distributed lag (ARDL) model with nonstationary variables:

$$y_t = \delta + \theta_1 y_{t-1} + \delta_0 x_t + \delta_1 x_{t-1} + v_t$$

- If y and x are cointegrated, it means that there is a long-run relationship between them
- To derive this exact relationship, we set

$y_t = y_{t-1} = y$, $x_t = x_{t-1} = x$ and $v_t = 0$

$$y(1-\theta_1) = \delta + (\delta_0 + \delta_1)x$$

$$y = \beta_1 + \beta_2 x$$

SSRMC

Cointegration

- Error Correction Model (Cont'd)
- Re-arrange the ARDL model by adding the term $-y_{t-1}$ to both sides of the equation, and adding the term $-\delta_0 x_{t-1} + \delta_0 x_{t-1}$ to the RHS, we obtain

$$\Delta y_t = \delta + (\theta_1 - 1) y_{t-1} + \delta_0 (x_t - x_{t-1}) + (\delta_0 + \delta_1) x_{t-1} + v_t$$

- Manipulating this we get

$$\Delta y_t = (\theta_1 - 1)\left(\frac{\delta}{(\theta_1 - 1)} + y_{t-1} + \frac{(\delta_0 + \delta_1)}{(\theta_1 - 1)} x_{t-1} \right) + \delta_0 \Delta x_t + v_t$$

$$\Delta y_t = -\alpha (y_{t-1} - \beta_1 - \beta_2 x_{t-1}) + \delta_0 \Delta x_t + v_t$$

- This is called an Error Correction Model

Cointegration

- Error Correction Model example:
- The relation between federal funds rate (F_t) and three-year bond rate (B_t)

$$\Delta\hat{B}_t = -0.142\left(B_{t-1} - 1.429 - 0.777F_{t-1}\right) + 0.842\Delta F_t - 0.327\Delta F_{t-1}$$
$$(t) \quad (2.857) \qquad\qquad (9.387) \quad (3.855)$$

$$\hat{e}_{t-1} = \left(B_{t-1} - 1.429 - 0.777F_{t-1}\right)$$

- The ADF test for stationarity is

$$\Delta\hat{e}_t = -0.169\hat{e}_{t-1} + 0.180\Delta\hat{e}_{t-1}$$
$$(t) \quad (-3.929)$$

Regression When There Is No Cointegration

- We need to convert nonstationary, non-cointegrated series to stationary series
- The technique depends on whether the variables are
  - – difference stationary ($y_t = \alpha + y_{t-1} + v_t$) or
  - – trend stationary ($y_t = \alpha + \delta t + v_t$)
- Difference stationary: first differencing

$$\Delta y_t = \alpha + v_t$$

- Trend stationary: de-trending

$$y_t - \alpha - \delta t = v_t$$

Summary

- If variables are stationary, or I(1) and cointegrated, we can estimate a regression relationship between the levels of those variables without fear of encountering a spurious regression
- If the variables are I(1) and not cointegrated, we need to estimate a relationship in first differences, with or without the constant term
- If they are trend stationary, we can either de-trend the series first and then perform regression analysis with the stationary (de-trended) variables or, alternatively, estimate a regression relationship that includes a trend variable

Summary

- Stationary and Nonstationary Variables
- Spurious Regressions
- Unit Root Tests for Nonstationarity
- Cointegration
- Regression When There Is No Cointegration